Choosing Your Beliefs

Guido Boella[1], Célia da Costa Pereira[2], Gabriella Pigozzi[3], Andrea Tettamanzi[2], and Leendert van der Torre[3]

[1] Università di Torino, Dipartimento di Informatica
10149, Torino, Cso Svizzera 185, Italia
guido@di.unito.it
[2] Università degli Studi di Milano, Dipartimento di Tecnologie dell'Informazione
26013, Crema, via Bramante 65, Italy
pereira@dti.unimi.it, andrea.tettamanzi@unimi.it
[3] Université du Luxembourg, Computer Science and Communication
L-1359, Luxembourg, rue Richard Coudenhove - Kalergi 6, Luxembourg
gabriella.pigozzi@uni.lu, leendert@vandertorre.com

Abstract This paper presents and discusses a novel approach to indeterministic belief revision. An indeterministic belief revision operator assumes that, when an agent is confronted with a new piece of information, it can revise its belief sets in more than one way. We define a rational agent not only in terms of what it believes but also of what it desires and wants to achieve. Hence, we propose that the agent's goals play a role in the choice of (possibly) one of the several available revision options. Properties of the new belief revision mechanism are also investigated.

Keywords. Rational agents, indeterministic belief revision, qualitative decision theory.

1 Motivating example

Suppose that you believe that:

1. A liberal policy leads to decrease of unemployment, and
2. A decrease of unemployment leads to re-election,

and you desire to be re-elected. Therefore you execute a plan based on a liberal policy, or do something else to decrease unemployment, and secure your re-election.

Now suppose that someone informs you that a liberal policy does not lead to re-election. Assume the person telling you this is very trustworthy, has a good reputation, so you believe what he is telling you. This implies that the three beliefs cannot hold together, and you have to give up one of them.

Now assume in addition that when you give up your belief in the first rule, you still have another plan to decrease unemployment, and thus to be re-elected, for example by increasing government spending on public works like building

Dagstuhl Seminar Proceedings 07122
Normative Multi-agent Systems
http://drops.dagstuhl.de/opus/volltexte/2007/938

bridges. However, if you give up your second belief that lower unemployment leads to re-election, you do not have an alternative plan to achieve re-election. In that case, you might reason by cases as follows.

1. Let us first assume that the first belief that liberal policy leads to lower unemployment is factually wrong, whereas the second belief is true. If you choose to retain the first (wrong) belief and to reject the second one, then you will do nothing and you will not succeed in being re-elected. But, had you kept your belief in the second rule and rejected the first belief, you could have increased public spending in order to decrease unemployment, and therefore you could have achieved your goal to be re-elected. To conclude, choosing to maintain the first belief, you risk to miss a goal you could have achieved.
2. Let us now assume assume that the first belief is actually true and the second belief is wrong. If you choose to keep the second (wrong) belief that decreasing unemployment leads to re-election, you will increase public spending, but you will not achieve the goal of being re-elected. However, even if you had chosen the right revision, i.e. to retain the first belief and reject the second one, you could have not achieved your goal of re-election. To conclude, by choosing the second (wrong) rule, you believed you could achieve a goal when you could not, so you will be disappointed for trying in vain but at least you tried to reach your aim.

The moral of the story is that, if you are interested only in realizing your goal (and there are no other goals relevant for you), then choosing the second belief — *even when it is factually wrong* — is the only rational choice. This is because, independently of the second belief being right or wrong, by choosing that belief you will end up in an optimal state. Moreover, in one situation — the first one — you will end up in a better state if you choose the second belief than if you choose the first one. Summarizing, you should drop the first belief, because in that way, you keep open all possibilities to achieve your goal.

We can formalize the above example, by defining the following atomic propositions:

p you are following a liberal policy;
u unemployment decreases;
r you will be re-elected;
s you are increasing public spending.

The belief base before being told that a liberal policy does not lead to re-election ($\neg(p \supset r)$) would contain the three formulas $p \supset u$, $u \supset r$, and $s \supset u$. You desire, first of all, to be re-elected, r, and, if possible, not to increase public spending, $\neg s$. Adding $\neg(p \supset r)$ to your beliefs would make them inconsistent. Therefore, you have to revise your beliefs by giving up either $p \supset u$ or $u \supset r$. The choice you make may depend on the goals you can achieve in the alternatives: if you give up $p \supset u$, your plan will be to increase public spending, so you will not achieve $\neg s$, but might succeed in achieving r; if you give up $u \supset r$, your plan will be to do nothing, so you will certainly not achieve r, but you will fulfill $\neg s$.

Depending on the payoff you expect from r and $\neg s$, you could prefer one or the other alternative.

We use the re-election example as a running example throughout the paper. There are some particular issues involved in the re-election situation, such as temporal references. However, we believe that many problems can be phrased in a similar way, including examples referring to factual statements about the present state of the world rather than hypothetical statements referring to the future as in the re-election example. For example, p may stand for "it is a public holiday", q for "your favorite restaurant is open", r stands for "eating in the restaurant", and you are informed that the restaurant is closed on public holidays. In that case, you would give up your belief that it is a public holiday, because that is the only way to achieve your goal to eat in the restaurant. You may drive to the restaurant in vain, but it would be much worse to give up the belief that the restaurant is open, only to find out later that you could have eaten there. At least, as long as we assume that this line of reasoning does not interfere with other goals. For example, the goal not to drive to restaurants in vain should not be preferred to the goal to eat in the restaurant, and there should not be information that it is much more likely that it is a public holiday than that the restaurant is open, and so on.

The choice among belief sets is distinct from other decision problems, due to the possibility of wishful thinking. Consider for example that you desire that liberal policy leads to lower unemployment, and that this desire is preferred to the desire to be re-elected. What will you do? At least in a naive approach, you could reason by cases as follows. Assume that you choose the first belief, in that case you believe that will achieve the desire that liberal policy leads to lower unemployment. Assume that you choose the second belief, in that case you believe that you will achieve the goal to be re-elected. Since the first goal is more important than the second one, you choose the first belief. Analogously, in the restaurant example, if you like public holidays and this desire is stronger than your desire to eat in the restaurant, than a naive reasoner may choose the first belief set. However, this is again a case of wishful thinking, and not a valid reason not to go to the restaurant. Instead, one should reason by cases as follows. Either it is a public holiday or not. If it is a public holiday, it does not matter what we do, we always achieve the goal that it is a public holiday and we never achieve the goal to go to the restaurant, since it is closed. If it is not a public holiday, then we will never achieve the desire for public holidays, but we may achieve the desire to eat in the restaurant — at least, when we go there. Our formal framework illustrates why the line of reasoning leading to the decision not to go to the restaurant is fallacious.

The idea of this paper is inspired by the notion of *conventional wisdom* (CW) as introduced by economist John Kenneth Galbraith:

> We associate truth with convenience, with what most closely accords with self-interest and personal well-being. ([14], p. 34)

That is, CW consists of "ideas that are convenient, appealing". This is the rationale for keeping them. One basic brick of CW could then be the fact that

some ideas are maintained because they maximize the goals that the agents (believe) they can achieve. This work may be seen as an initial attempt to formally capture the concept of a CW agent. In the following we provide a logical framework that models how a CW agent revises its beliefs.

The paper is structured as follows. In Section 2 we introduce the aim of this paper, the used methodology and particular challenges encountered. In Section 3 we introduce the agent theory we use in our approach, and in Section 4 we introduce an indeterministic belief change operator in this agent theory. In Section 5 we define the choice among beliefs as a decision problem in the agent theory. We conclude the paper by clarifying the relation of our proposal with some existing work (Section 6) and with some final remarks and ideas for future research (Section 7).

2 Aim, methodology and challenges

The research problem of this paper is to develop a formal model to reason about the kind of choices among belief sets discussed in the previous section, and to generalize the example above in case of additional beliefs, multiple goals with preferences among them, conditional desires, a way to take violated goals into account, and so on.

We use a combination of the framework of belief revision together with a qualitative decision theory. Classical approaches to belief revision assume that, when an agent revises its belief set in view of a new input, the outcome is well-determined. This picture, however, is not realistic. When an agent revises its beliefs in the light of some new fact, it often has more than one available alternative. Approaches to belief revision that do not stipulate the existence of a single revision option are called *indeterministic* [16,19]. In this paper we suggest that one possible policy an agent can use in order to choose among available alternatives is to check the effect of the different revisions on the agent's set of goals.

Moreover, for the qualitative decision theory we are inspired by agent theories such as the BOID architecture [2], the framework of goal generation in 3APL as developed by van Riemsdijk and colleagues [20], and [5]. In particular, our agent model is based on one of the versions of 3APL, because the belief base in the mental state of a 3APL agent is a consistent set of propositional sentences, just like in the framework of belief revision. However, we do not include "planning rules" or "practical reasoning rules" representing which action to choose in a particular state, because we aim for a modular agent architecture. We assume that there is a planning module, which would take a set of goals, actions, and an initial world state representation in input and produce a solution plan in output. This planning module might rely on the well-known graphplan algorithm, or any other AI planner: as in object-oriented programming, we *encapsulate* the planner within a well-defined interface and overlook the implementation details of how a solution plan is found. This is in line with the BOID architecture [2], where the planning component is kept separate from the remainder of agent deliberation.

In other words, we model the choice among belief sets essentially as a decision problem, that is, as a choice among a set of alternatives. We do not use classical decision theory (utility function, probability distribution, and the decision rule to maximize expected utility), but a qualitative version based on maximizing achieved goals and minimizing violated goals in an abstract agent theory (see e.g. [8] for various approaches to formalize the decision process of what an agent should do), because such qualitative decision theories include beliefs and therefore are easier to combine with the theory of belief revision. However, what precisely are the alternatives?

An indeterministic belief revision operator associates multiple revision options to a belief set that turns out to be inconsistent as a consequence of a new piece of information. Our revision mechanism selects the revision alternative that allows the agent to maximize its achievable goals. However, it will not always be possible to select exactly one revision alternative. For example, there may be one preferred goal set but two revision alternatives that lead the agent to achieve it. In this case, the two belief revision candidates are said to be equivalent. In Section 5.3 we will provide conditions under which a revision for a CW agent is deterministic, that is, when our revision operator can select exactly one revision alternative.

Besides the issue of wishful thinking, another complicating factor when choosing among belief sets in the context of conditional desire rules, is that a maximization of goals may lead to a meta-goal to derive goals. However, deriving goals by itself does not have to be desirable. In contrast, it may even be argued that fewer goals is better than more goals, as you risk to violate goals and become unhappy (as in Buddhism). One possible solution would be taking also goal violations into account. However, we do not address this issue in this paper.

3 An abstract agent theory

In this section, we represent the formalism which is used throughout the paper.

3.1 A brief introduction to AI planning and agent theory

Any agent, be it biological or artificial, must possess knowledge of the environment it operates in, in the form of e.g. *beliefs*. Furthermore, a necessary condition for an entity to be an *agent* is that it acts. We shall call the factors that motivate an agent to act *desires*. For artificial agents, desires may be the purposes an agent was created for.

Desires are necessary, not sufficient, conditions for action. When a desire is met by other conditions that make it possible for an agent to act, that desire becomes a *goal*.

The reasoning side of acting is known as practical reasoning or deliberation, which may include *planning*. Planning is a process that chooses and organizes actions by anticipating their expected effects with the purpose of achieving as good as possible some pre-stated objectives or goals.

Acting does not always imply planning. An agent deliberates or "chooses" to plan when it has to make difficult or unusual tasks; or when there are high risks or interests; or when there is the necessity of synchronizing several tasks which are part of a dynamic system. Otherwise, its acting may be based on simple stimulus-response rules.

Formally, AI planning may be defined as follows. Disposing of

(i) a representation of the initial state of the world I;
(ii) a set A of actions; and
(iii) a description of the goals to be reached G;

planning consists of finding a sequence of actions from A — *plan* — whose execution, from the initial state I, leads to a final state in which G is satisfied. Each action has a set of conditions it needs to be executed — *preconditions*— and once executed, it produces one or more *effects* which changes the world from a state to another. One of the theoretical motivations for planning is its utilization as a component of the rational behavior of an agent. In this case, this is all about providing the agent which has the task of constructing and/or executing a plan, reasoning capabilities like being able to react according to its perceptions, i.e., according to the changes in its mental state (beliefs and desires). In this case, planning consists in constructing a plan whose execution leads from the initial state to a final state which satisfies the *goals emerged during the planning process or during the plan execution.*

In the example, we may assume you dispose of two actions:

a_p (implement a liberal policy), with no precondition, and with effect p;
a_s (increase public spending), with no precondition, and with effect s.

Our formalism is inspired by one of the variants of the agent programming language 3APL used in [20]. However, unlike [20], the objective of our formalism is to analyze, not to develop, agent systems. More precisely, our agent must single out the *best* set of goals to be given as an input to a traditional planner. That is because the intentions of the agent are not considered. We merely consider beliefs (knowledge the agent has about the world states), desires (or motivations) and relations (desire-adopting rules) defining how the desire base will change with the acquisition of new beliefs and/or new desires. The goal generation process that underlies this work is very much in line with the work carried out in [21] on *oversubscription planning problems*, in which the main objective is to find the maximal set of desires to be reached in a given period and with a limited quantity of resources, and with goal generation in the BOID architecture [2].

3.2 Beliefs, Desires, and Goals

The basic components of our language are *beliefs* and *desires*. Beliefs are represented by means of a *belief base*. A belief base is a finite and consistent set of propositional formulas describing the information the agent has about the world and internal information. Desires are represented by means of a *desire base*. A

desire base consists of a set of propositional formulas which represent the situations the agent would like to achieve. However, unlike the belief base, a desire base may be inconsistent, e.g., $\{p, \neg p\}$.

Definition 1 (Belief Base Σ and Desire Base Γ) *Let $\mathcal{L}$ be a propositional language with $\top$ a tautology, and the logical connectives $\wedge$ and $\neg$ with the usual meaning. The agent's belief base Σ is a consistent finite set of atomic propositions like ϕ, φ, ψ, ... and compound propositions like $\neg\phi$, $\varphi \wedge \psi$, and so on. Σ can also be represented as the conjunction of its propositional formulas. The agent's desire base is a possibly inconsistent finite set of sentences denoted by Γ, with $\Gamma \subseteq \mathcal{L}$.*

We use modal languages to talk about the belief and desire bases of the agent. Since the belief and desire bases of an agent are completely separated, there is no need to nest the operators **B** and **D**.

Definition 2 (Belief Formulas β and Desire Formulas κ) *Given any formula ϕ of $\mathcal{L}$, $\mathbf{B}\phi$ means that ϕ is believed whereas $\mathbf{D}\phi$ means that ϕ is desired. The languages of belief formulas $\beta \in \mathcal{L}_B$ and desire formulas $\kappa \in \mathcal{L}_D$ are defined as follows:*

$$\beta ::= \top | \mathbf{B}\phi | \neg\beta | \beta_1 \wedge \beta_2$$

$$\kappa ::= \top | \mathbf{D}\phi | \neg\kappa | \kappa_1 \wedge \kappa_2$$

Following van Riemsdijk and colleagues, the antecedent of a desire-adoption rule consists of a belief condition and a desire condition; the consequent is a propositional formula. Intuitively, this means that if the belief and the desire conditions in the antecedent hold, the formula in the consequent is automatically adopted as a desire. Note that this implies that in the antecedent we may have for example a disjunction of two beliefs or a disjunction of two desires, but we cannot have a disjunction of a belief and a desire.

Definition 3 (Desire-Adoption Rules $\mathcal{R}_D$) *The language of desire-adoption rules $\mathcal{L}_R$ is defined as follows:*

$$\mathcal{L}_R = \{\beta, \kappa \Rightarrow_D^+ \phi \mid \beta \in \mathcal{L}_B, \kappa \in \mathcal{L}_D, \phi \in \mathcal{L}\}$$

The set of desire-adoption rules $\mathcal{R}_D$ of an agent is a finite subset of $\mathcal{L}_R$.

Goals, in contrast to desires, are represented by consistent desire bases. There are various ways to generate candidate goal sets from the desire adoption rules, as discussed in the remainder of this section.

Definition 4 (Candidate Goal Set Γ^*) *A candidate goal set Γ^* is a consistent subset of Γ.*

3.3 Mental State Representation

We assume that an agent is equipped with three bases:

- belief base $\Sigma \subseteq \mathcal{L}$;
- desire base: $\Gamma \subseteq \mathcal{L}$;
- desire-adoption rule base $\mathcal{R}_D$;

The state of an agent is completely described by a triple $\mathcal{S} = \langle \Sigma, \Gamma, \mathcal{R}_D \rangle$. In addition, we assume that each agent can be described using a $\mathcal{P}$-dependent function $\mathcal{F}_\mathcal{P}$, a pay-off function $f : \mathcal{L} \to \mathbb{R}$, a goal selection function G, and a belief revision operator $*$, as discussed below.

In the example,

$$\begin{aligned} \Sigma &= \{\neg(p \wedge \neg u), \neg(u \wedge \neg r), \neg(s \wedge \neg u)\}, \\ \Gamma &= \{r, \neg s\}, \\ \mathcal{R}_D &= \{\top, \top \Rightarrow^+_D r; \top, \top \Rightarrow^+_D \neg s\}. \end{aligned}$$

The semantics we adopt for the belief formulas is standard.

Definition 5 (Semantics of Belief Formulas) *Let $\phi \in \mathcal{L}$, $\beta \in \mathcal{L}_B$, and let $\langle \Sigma, \Gamma, \mathcal{R}_D \rangle$ be the mental state of an agent. The semantics of belief formulas is given as*

$$\begin{aligned} &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \top \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \mathbf{B}\phi \Leftrightarrow \Sigma \models \phi \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \neg\beta \Leftrightarrow \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \not\models_{\mathcal{L}_B} \beta \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \beta_1 \wedge \beta_2 \Leftrightarrow \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \beta_1 \text{ and } \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_B} \beta_2 \end{aligned}$$

The semantics we adopt for desire formulas is similar to the semantics of goal formulas proposed in [20].

Definition 6 (Semantics of Desire Formulas) *Let $\phi \in \mathcal{L}$, $\kappa \in \mathcal{L}_D$, and let $\langle \Sigma, \Gamma, \mathcal{R}_D \rangle$ be the mental state of an agent. The semantics of desire formulas is given as*

$$\begin{aligned} &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \top \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \mathbf{D}\phi \Leftrightarrow \exists \Gamma' \subseteq \Gamma : (\Gamma' \not\models \bot \text{ and } \Gamma' \models \phi) \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \neg\kappa \Leftrightarrow \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \not\models_{\mathcal{L}_D} \kappa \\ &\langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \kappa_1 \wedge \kappa_2 \Leftrightarrow \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \kappa_1 \text{ and } \langle \Sigma, \Gamma, \mathcal{R}_D \rangle \models_{\mathcal{L}_D} \kappa_2 \end{aligned}$$

We expect a rational agent to try and manipulate its surrounding environment to fulfill its goals. In general, given a planning problem $\mathcal{P}$, not all goals can be fulfilled. For example, if in the description of the problem there is no action whose list of effects includes a goal, that goal will not be feasible. Hence, we assume a $\mathcal{P}$-dependent function $\mathcal{F}_\mathcal{P}$ that, given a belief base Σ and a goal set Γ^*, returns $\top$ if Γ^* is feasible and $\bot$ otherwise.

Function $\mathcal{F}_\mathcal{P}$ obeys the following axioms:

1. for all Σ, $\mathcal{F}_{\mathcal{P}}(\Sigma, \emptyset) = \top$ (an empty set of goals is always feasible);
2. for all Σ, Γ_1^*, Γ_2^*, if $\Gamma_1^* \subseteq \Gamma_2^*$,

$$\mathcal{F}_{\mathcal{P}}(\Sigma, \Gamma_1^*) = \bot \Rightarrow \mathcal{F}_{\mathcal{P}}(\Sigma, \Gamma_2^*) = \bot,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma, \Gamma_2^*) = \top \Rightarrow \mathcal{F}_{\mathcal{P}}(\Sigma, \Gamma_1^*) = \top$$

(goal set feasibility is monotonous).

In the example, where $\Sigma = \{\neg(p \wedge \neg u), \neg(u \wedge \neg r), \neg(s \wedge \neg u)\}$,

$$\mathcal{F}_{\mathcal{P}}(\Sigma, \emptyset) = \top,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma, \{r\}) = \top,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma, \{\neg s\}) = \top,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma, \{r, \neg s\}) = \top;$$

however, if we had $\Sigma' = \{p \wedge \neg r, \neg(u \wedge \neg r), \neg(s \wedge \neg u)\}$,

$$\mathcal{F}_{\mathcal{P}}(\Sigma', \emptyset) = \top,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma', \{r\}) = \bot,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma', \{\neg s\}) = \top,$$
$$\mathcal{F}_{\mathcal{P}}(\Sigma', \{r, \neg s\}) = \bot.$$

Definition 7 (Feasible Candidate Goal Set) *A candidate goal set Γ^* is feasible for a planning problem $\mathcal{P}$ if and only if $\mathcal{F}_{\mathcal{P}}(\Sigma, \Gamma^*) = \top$.*

3.4 Comparing Candidate Goals and Sets of Candidate Goals

In this section we define one possible way in which an agent can choose among different sets of candidate goals. The particular choice made in this section is meant to illustrate goal comparison in the agent theory. If this particular way is replaced by another one, then still the general problem on choosing beliefs holds, and our solution can be applied.

¿From a desire base Γ, several candidate goal sets Γ_i^*, $1 \leq i \leq n$, may be derived. How can an agent choose among all the possible Γ_i^*? It is unrealistic to assume that for a rational agent all goals have the same importance. Therefore, we use the notion of expected pay-off to represent how relevant each goal is for the agent. The idea is that a rational agent tries to choose a set of goals which, first of all, is feasible and, secondly, gives the highest pay-off.

A pay-off function is a function $f : \mathcal{L} \rightarrow \mathbb{R}$ which associates a real value (the pay-off) to every formula in $\mathcal{L}$. Given $\phi \in \mathcal{L}$, $f(\phi)$ is the pay-off the agent would receive if ϕ were true. For all $\phi \in \Gamma$, we assume that $f(\phi) > 0$. In other words, a rational agent cannot desire something that, if realized, would bring no benefit.

One problem with pay-offs is that an agent may not always be able to attach a precise numerical value to its desires. An alternative approach would be to assume a total order over an agent's desires. In either case, we can define a total

order $\succeq$ between goals, such that, for all $\phi_1, \phi_2 \in \mathcal{L}$, $\phi_1 \succeq \phi_2$ iff $f(\phi_1) \geq f(\phi_2)$, or the agent desires ϕ_1 at least as much as it desires ϕ_2.

In the example, we could express the fact you want most of all to be reelected and, if possible, would rather not increase public spending, by defining $f(r) > f(\neg s)$, e.g., $f(r) = 100$, $f(\neg s) = 10$. In case you find it unnatural to assign arbitrary numerical values to these payoffs, you could just use a total order $\succeq$, and define $r \succ \neg s$.

The $\succeq$ relation can be extended from candidate goals to sets of candidate goals. For the qualitative ordering, we have that a goal set Γ_1^* is preferred to another one Γ_2^* if, considering only the goals occurring in one of the sets, the best goals are in Γ_1^* or the worst goals are in Γ_2^*. Note that $\succeq$ is connected and therefore a total pre-order, i.e., we always have $\Gamma_1^* \succeq \Gamma_2^*$ or $\Gamma_2^* \succeq \Gamma_1^*$.

Definition 8 (Preference between Sets of Candidate Goals) *Given two candidate goal sets Γ_1^* and Γ_2^*:*

- *We say that Γ_1^* is at least as preferred as Γ_2^* (denoted by $\Gamma_1^* \succeq \Gamma_2^*$):*

$$\sum_{\phi \in \Gamma_1^*} f(\phi) \geq \sum_{\phi \in \Gamma_2^*} f(\phi)$$

  if the pay-offs are defined.
- *If a preference relation over candidate goals is given, let $\Gamma_1' = \Gamma_1^* \setminus \Gamma_2^*$ and $\Gamma_2' = \Gamma_2^* \setminus \Gamma_1^*$. We then say that $\Gamma_1^* \succeq \Gamma_2^*$ iff one of the following two conditions is satisfied:*
  1. *$\forall \phi_2 \in \Gamma_2'$, $\exists \phi_1 \in \Gamma_1'$, s.t. $\phi_1 \succeq \phi_2$;*
  2. *$\forall \phi_1 \in \Gamma_1'$, $\exists \phi_2 \in \Gamma_2'$, such that $\phi_1 \succeq \phi_2$.*

In the example, it is easy to verify that the following relation holds in either cases (if a payoff function is defined or if a preference order is used):

$$\{r, \neg s\} \succ \{r\} \succ \{\neg s\} \succ \emptyset.$$

3.5 Defining the Goal Set Selection Function

In general, given a set of desires Γ, there may be many possible candidate goal sets. A rational agent in state $\mathcal{S} = \langle \Sigma, \Gamma, \mathcal{R}_D \rangle$ will select as the set of goals it wants to pursue one precise candidate goal set Γ^* among the most preferred feasible candidate goal sets, which depends on $\mathcal{S}$.

Let us call G the function which maps a state $\mathcal{S}$ into the goal set selected by a rational agent in state $\mathcal{S}$: $\Gamma^* = G(\mathcal{S})$.

In the example above, $G(\mathcal{S}) = \{r, \neg s\}$, because $\{r, \neg s\}$ is the most preferrable among the feasible goal sets.

4 Situating the Problem: Indeterministic Belief Change

> "Most models of belief change are deterministic in the sense that given a belief set and an input, the resulting belief set is well-determined. There is no scope for chance in selecting the new belief set. Clearly, this is not a realistic feature, but it makes the models much simpler and easier to handle, not least from a computational point of view. In indeterministic belief change, the subjection of a specified belief set to a specified input has more than one admissible outcome.
> Indeterministic operators can be constructed as sets of deterministic operations. Hence, given n deterministic revision operators $*_1, *_2, \ldots, *_n$, $* = \{*_1, *_2, \ldots, *_n\}$ can be used as an indeterministic operator." [15]

Let us consider a belief set Σ and a new belief β. The revision of Σ in light of a new belief β is simply:

$$\Sigma * \beta \in \{\Sigma *_1 \beta, \Sigma *_2 \beta, \ldots \Sigma *_n \beta\}. \tag{1}$$

More precisely, revising the belief set Σ with the indeterministic operator $*$ in light of new belief β leads to one of the n belief revision results:

$$\Sigma * \beta \in \{\Sigma^1_\beta, \Sigma^2_\beta, \ldots \Sigma^n_\beta\}, \tag{2}$$

where Σ^i_β is the ith possible belief revision result.

Applying operator $*$ is then equivalent to applying one of the virtual operators $*_i$ contained in its definition. While the rationality of an agent does not suggest any criterion to prefer one revision over the others, a defining feature of a CW agent is that it will choose which revision to adopt based on the consequence of that choice. One important consequence is the set of goals the agent will decide to pursue.

In the example, $\beta = \mathbf{B}(p \wedge \neg r)$, and

$$\Sigma * \beta \in \left\{ \begin{array}{l} \Sigma^1_\beta = \{p \wedge \neg r, \neg(u \wedge \neg r), \neg(s \wedge \neg u)\}, \\ \Sigma^2_\beta = \{p \wedge \neg r, \neg(p \wedge \neg u), \neg(s \wedge \neg u)\} \end{array} \right\}. \tag{3}$$

In the next sections we propose some possible ways to tackle the problem of choosing one of the revision options among the different available.

5 Belief Revision as a Decision Problem

By considering an indeterministic belief revision, we admit $\Sigma * \beta$ to have more than one possible result. In this case, the agent must select one among all possible revisions. Many criteria can be considered for selection. One of the criteria is to choose the belief revision operator for which the goal set selection function returns the most preferrable goal set. In other words, selecting the revision amounts to solve an optimization problem.

5.1 Indeterministic State Change

The indeterminism of belief revision influences the desire-updating process. In fact, the belief revision operator is just a part of the state-change operator, which is indeterministic as well, as a consequence of the indeterminism of belief revision. Therefore,

$$\mathcal{S}_\beta \in \{\mathcal{S}^1_\beta, \mathcal{S}^2_\beta, \ldots, \mathcal{S}^n_\beta\}, \tag{4}$$

where $\mathcal{S}^i_\beta = \langle \Sigma^i_\beta, \Gamma^i_\beta, \mathcal{R}_D \rangle$.

Which goal set is selected by an agent depends on G:

$$G(\mathcal{S}_\beta) \in \{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}. \tag{5}$$

In the example,

$$G(\mathcal{S}_\beta) \in \{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta)),$$

where $G(\mathcal{S}^1_\beta) = \{r\}$ and $G(\mathcal{S}^2_\beta) = \{\neg s\}$. The following table summarizes the possibilities you may face when choosing between the two alternative revisions.

reality $\rightarrow$ $\downarrow$ beliefs	$\not\models p \supset u$ $\models u \supset r$	$\models p \supset u$ $\not\models u \supset r$
Σ^1_β plan: increase public spending	r is achieved $\neg s$ is not achieved pay-off $= f(r)$	no desire is achieved pay-off $= 0$
Σ^2_β plan: do nothing	r is not achieved $\neg s$ is achieved pay-off $= f(\neg s)$	

A traditional rational agent could not choose one of the $G(\mathcal{S}^i_\beta)$ because they are incomparable. Now, for a CW agent,

$$G(\mathcal{S}_\beta) \in \mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}, \tag{6}$$

where $\mathrm{PS}(S)$ denotes the preferred set of S defined as follows:

Definition 9 (Preferred Set PS**)** *Given two sets S and X such that $S \subseteq X$, and given a preference relation $\succeq$ over X, the preferred set of S is*

$$\mathrm{PS}(S) = \{x \in S : \forall x' \in S, x \succeq x'\}. \tag{7}$$

Since in the example $G(\mathcal{S}^1_\beta) = \{r\}$ and $G(\mathcal{S}^2_\beta) = \{\neg s\}$, and $\{r\} \succ \{\neg s\}$,

$$\mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta)\} = \mathrm{PS}\{\{r\}, \{\neg s\}\} = \{\{r\}\}.$$

Therefore, a CW agent should choose revision Σ^1_β, because it is the only revision whereby you could possibly end up being re-elected, which is what you desire most. This is in agreement with the intuition underlying the motivating example.

5.2 Choosing a Revision

As long as different revisions lead to distinct goal sets with different degrees of preference, it is clear what revision a CW agent should choose. However, we can distinguish two situations in which the choice is less trivial:

- there is just one preferred goal set Γ^*, but more than one alternative options lead to Γ^*;
- there is no unique preferred goal set; that is, there are different goal sets $\Gamma^*_1, \ldots, \Gamma^*_m$, none of which is strictly preferred to the others, i.e., for all $i, j \in \{1, \ldots, m\}$, $\Gamma^*_i \succeq \Gamma^*_j$.

In these cases, some alternative belief revisions lead to equally preferred goal sets, and such revisions may be regarded as equivalent.

Definition 10 (Equivalence between Belief Revision Candidates) *A belief revision candidate Σ^1_β is equivalent to another belief revision candidate Σ^2_β (denoted by $\Sigma^1_\beta \approx \Sigma^2_\beta$), if and only if $G(\mathcal{S}^1_\beta) \succeq G(\mathcal{S}^2_\beta)$ and $G(\mathcal{S}^2_\beta) \succeq G(\mathcal{S}^1_\beta)$.*

It is easy to verify that $\approx$ is a standard equivalence relation, i.e., reflexive, symmetric, and transitive.

The choice of which revision outcome to adopt may thus be deterministic or indeterministic. It is indeterministic in the two cases presented above. More precisely, the choice depends on the preference relations over the goal sets, which determines the equivalence between revision candidates:

- if $\|\mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}\| = 1$, i.e., the equivalent class of a preferred belief revision is a singleton and, if there is no i, j such that $G(\mathcal{S}^i_\beta) = G(\mathcal{S}^j_\beta)$, the choice of the belief operator is obviously deterministic;
- if $\|\mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}\| = 1$, and there is at least a couple i, j such that $G(\mathcal{S}^i_\beta) = G(\mathcal{S}^j_\beta)$, the choice may be indeterministic, if two or more distinct revisions lead to one and the most preferred goal set, but also indifferent in practice.
- if $\|\mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}\| > 1$, the choice is indeterministic;

It is important to notice that an agent that has to choose between Σ^i_β and Σ^j_β which lead to the same goal set (as in the second case above) is in a different situation than an agent who has to randomly choose among a number of competing revisions (as in the third case above). In the second case, whatever the agent's choice is, the goals are the same; in the third case, depending on the agent's choice, the goals the agent will pursue may vary. In general, a random choice is hardly a rational option. But, when an agent is in the second situation, it knows that, no matter which revision it chooses, the outcome does not change. In such a context, a random choice becomes a rational option.

Proposition 1 *Let $*$ be an indeterministic belief operator, and n be the number of possible belief revisions candidate. We have:*

$$1 \leq \|\mathrm{PS}\{G(\mathcal{S}^1_\beta), G(\mathcal{S}^2_\beta), \ldots, G(\mathcal{S}^n_\beta)\}\| \leq n.$$

5.3 Conditions for Determinism of a CW Agent

Traditional indeterministic belief revision approaches allow for the result of belief revision to be indeterminate in the sense that there may be many possible revision alternatives that are equally rational. Our proposal builds on the idea that what an agent wishes to achieve can play a role in the choice of which beliefs to reject and which beliefs to retain. The example we have been using in this paper also tries to capture the intuition that an agent who behaves in this manner is rational. Our richer model can distinguish one revision alternative from the other depending on the effect that each option has on the agent's goal set. Hence, under certain conditions, the choice among several revision alternatives can be reduced to one. This is what we want to investigate now, that is we want to investigate the conditions under which a revision for a CW agent is deterministic even if an indeterministic revision operator is used, i.e., $\|\mathrm{PS}\{G(\mathcal{S}^i_\beta)\}_{i=1,\ldots}\| = 1$ and, for all i, j, $G(\mathcal{S}^i_\beta) \neq G(\mathcal{S}^j_\beta)$. Determinism may be desirable, for instance, in agent programming, in those cases where predictability of the agent's behaviour is a requirement.

Observation 1 *$\Sigma * \beta$ is deterministic in state $\mathcal{S} = \langle \Sigma, \Gamma, \mathcal{R}_D \rangle$, iff no two alternative revisions are equivalent, i.e., for all i, j, $\Sigma^i_\beta \not\approx \Sigma^j_\beta$.*

Proposition 2 *A sufficient condition for no two alternative revisions, Σ^i_β and Σ^j_β, being equivalent is that*

1. *for all i, j, $G(\mathcal{S}^i_\beta) \neq G(\mathcal{S}^j_\beta)$;*
2. *(a) if pay-offs are defined, for all i, j, $\sum_{\phi \in G(\mathcal{S}^i_\beta)} f(\phi) \neq \sum_{\phi \in G(\mathcal{S}^j_\beta)} f(\phi)$;*
   (b) if pay-offs are not defined, the preference relation on goals is strict, i.e., for all $\phi, \phi' \in G(\mathcal{S}_\beta)$, $\phi \neq \phi'$, $\phi \succeq \phi' \Rightarrow \phi' \not\succeq \phi$.

Proof: If pay-offs are defined, from Hypothesis 1 and 2a, by applying Definition 8, we obtain that either $(G(\mathcal{S}^i_\beta) \succeq G(\mathcal{S}^j_\beta)$ and $G(\mathcal{S}^j_\beta) \not\succeq G(\mathcal{S}^i_\beta))$ or $(G(\mathcal{S}^j_\beta) \succeq G(\mathcal{S}^i_\beta)$ and $G(\mathcal{S}^i_\beta) \not\succeq G(\mathcal{S}^j_\beta))$. Therefore, $\Sigma^i_\beta \not\approx \Sigma^j_\beta$.

If pay-offs are not defined, from Hypothesis 1 and 2b, by applying Definition 8, we obtain again $\Sigma^i_\beta \not\approx \Sigma^j_\beta$.

Therefore, no two alternative revisions can be equivalent. □

Proposition 3 *If the pay-off function f is such that, for all $\phi \in \mathcal{L}$, for all $\Psi \subseteq \mathcal{L} \setminus \{\phi\}$,*

$$f(\phi) \neq \sum_{\psi \in \Psi} f(\psi), \tag{8}$$

Condition 2a of Proposition 2 always holds.

The proof is trivial.

One might wonder how difficult it is to design a pay-off function that satisfies Inequality 8. The answer is, quite easy.

Proposition 4 *Given a rational, injective pay-off function f, there exists another pay-off function $\hat{f}$ such that*

1. *$\hat{f}$ satisfies Inequality 8;*
2. *for any desired δ, for all $\phi \in \mathcal{L}$, $|f(\phi) - \hat{f}(\phi)| < \delta$;*
3. *for all $\phi, \psi \in \mathcal{L}$, $f(\phi) > f(\psi) \Leftrightarrow \hat{f}(\phi) > \hat{f}(\psi)$.*

Proof: Since f is rational, there exists $u \in \mathbb{R}$ such that, for all $\phi \in \mathcal{L}$, $f(\phi) = n$ for some integer n. Let

$$\epsilon_0 < \min\{\delta, u, \min_{\phi \in \mathcal{L}} f(\phi)\}.$$

We define a sequence $\{\epsilon_i\}_{i=0,\ldots}$ such that $\epsilon_{i+1} = \epsilon_i/2$. It easy to verify that no element ϵ_i can be obtained as a sum of a finite number of other elements ϵ_j, with $i \neq j$. Now, let ϕ_1, ϕ_2, ... an effective enumeration of all formulas in $\mathcal{L}$ (such enumeration, which needs not be finite, exists for all recursively enumerable languages); pay-off function $\hat{f}$ may be defined, for all $i = 1, 2, \ldots$, as

$$\hat{f}(\phi_i) = f(\phi_i) + \epsilon_i.$$

Function $\hat{f}$ satisfies the three conditions of the thesis. □

6 Related work

6.1 Goal Change

In this paper we do not explain the process of goal generation and revision, i.e., we are not interested in how new goals arise in the light of new beliefs or desires. That aspect is considered, for example, in [5,6], where an approach has been proposed to dynamically construct the goal set to be pursued by a rational agent, by considering changes in its mental state. More precisely, the autors propose a general framework based on classical propositional logic, to represent changes in the mental state of the agent after the acquisition of new information and/or after the arising of new desires.

An important point of this framework, which distinguishes it from the framework used in this paper, is that the two aspects of how goals are selected by an agent and how the selected goals are achieved are not conceptually separated: this means, the goal selection mechanics depend on the planning process and then interactions between these two aspects are a part of the goal generation/revision process.

6.2 BOID

The BOID architecture [2] extends a classical planner with a component for goal generation. In this goal generation component, there are subcomponents for beliefs, obligations, intentions and desires [4]. The interaction among these

subcomponents is studied using a qualitative decision theory [3,12] and qualitative game theory [10] based on extensions of input/output logic [17,18,1]. using merging operators [9], as an extension of the 3APL programming language [11], and using defeasible logic [7]. Though in all of these approaches the relation between beliefs and goals plays a central role, in these papers the impact of goals on the choice among belief sets has not been studied.

6.3 Preference over Beliefs

Doyle suggests to have a preference order over belief sets [13]. We have however an indirect link from belief sets to feasible goals, and a preference order over these goals; and from these preferences over goals, we again derive the preferences over belief sets. Therefore, if one wanted to accept Doyle's suggestion, our work could be regarded as a method for deriving a rationally justified preference order over belief sets.

7 Conclusions

We have presented some preliminary ideas for a new approach aiming at resolving indeterminism in belief revision. The framework has been inspired by the concept of *conventional wisdom*, introduced by economist John Kenneth Galbraith. Revising a belief base with an indeterministic operator in light of a new belief leads to more than one possible revisions. In this case, a traditional rational agent would not be able to choose among the possible revision candidates. The idea we started to develop is that the agent, in this case, may evaluate the effects that the different revision options have on its goals. Therefore, it could choose a revision which maximizes its goals. In other words, selecting the revision would amount to solving an optimization problem. Finally, fundamental definitions and properties of the belief revision mechanisms have been given.

Some topics for further research:

1. We would like to add a function $V(S, G)$ that returns for a state together with the feasible goals, the goals which also have been generated but which are not feasible. Then we can define preferences not only over feasible goals as now, but also over unfeasible ones.
   Before doing so, we have to be clear about what we think goals are: achievement goals on events/punctual or maintenance goals on states/continuant? In other words: if goals are events, then once you achieve them, you can forget about them, like when you shoot at a target and you hit it; but if goals can be states, like "staying alive", this does not hold anymore: it isn't because you believe you're alive that you don't want to stay alive anymore.
2. Assume that we have conflict between $\mathbf{B}p$ and $\mathbf{B}\neg p$, and we can choose either one of these beliefs, or none at all. What can we say about this situation? Should we be more adventurous by believing either $\mathbf{B}p$ or $\mathbf{B}\neg p$ rather than believing nothing, and if so, under which conditions? Consider for example

the principle of goal generation that more information about beliefs leads to more goals (monotony in beliefs for practical reasoning rules). In that case, under suitable conditions, we can probably prove that we can ignore the choice in which we do not believe anything.
3. Another open point of this work concerns the situation in which even if we you consider preferences among goal sets, this would not be enough for determining the belief revision to be adopted – belief revision process remains indeterministic. In this case, it is necessary to provide a framework which deals with this situation. One possibility would be to keep revision options open waiting for a new input which help us to choose the more convenient revision.
4. We can formalize the present model as an abstraction of a more general model of decision making (e.g. taking inspiration from decision trees, Savage style decision theory, action logics etc.) and consider the rationality of our CW agent in this more general theory.

References

1. Boella, G., Hulstijn, J. and van der Torre, L., "Interaction in Normative Multi-Agent Systems", *Electronic Notes in Theoretical Computer Science*, **141(5)**, 2005, 135–162.
2. Broersen, J., Dastani, M., Hulstijn, J. and van der Torre, L., "Goal Generation in the BOID Architecture", *Cognitive Science Quarterly Journal*, **2**(3–4), 2002, 428–447.
3. Broersen, J., Dastani, M. and van der Torre, L., "Realistic Desires", *Journal of Applied Non-Classical Logics*, **12(2)**, 2002, 287–308.
4. Broersen, J., Dastani, M. and van der Torre, L., "Beliefs, Obligations, Intentions and Desires as components in an agent architecture", *International Journal of Intelligent Systems*, **20:9**, 2005, 893–919.
5. da Costa Pereira, C. and Tettamanzi, A., "Towards a Framework for Goal Revision", in: Pierre-Yves Schobbens, W. V. and Schwanen, G. (eds.), *BNAIC-06, Proceedings of the 18th Belgium-Netherlands Conference on Artificial Intelligence*, Namur, Belgium: University of Namur, 2006, 99–106.
6. da Costa Pereira, C., Tettamanzi, A. and Amgoud, L., "Goal Revision for a Rational Agent", in: Brewka, G., Coradeschi, S., Perini, A. and Traverso, P. (eds.), *ECAI 2006, Proceedings of the 17th European Conference on Artificial Intelligence*, Riva del Garda, Italy: IOS Press, 2006, 747–748.
7. Dastani, M., Governatori, G., Rotolo, A. and van der Torre, L., "Programming Cognitive Agents in Defeasible Logic", in: *Proceedings LPAR'05*, LNCS, Springer, 2005.
8. Dastani, M., Hulstijn, J. and van der Torre, L., "How to decide what to do?", *European Journal of Operational Research*, **160**(3), 2005, 762–784.
9. Dastani, M. and van der Torre, L., "Specifying the Merging of Desires into Goals in the Context of Beliefs", in: *Proceedings of The First Eurasian Conference on Advances in Information and Communication Technology (EurAsia ICT 2002)*, LNCS 2510, Springer, 2002, 824–831.
10. Dastani, M. and van der Torre, L., "Games for Cognitive Agents", in: *Proceedings of JELIA04*, LNAI 3229, 2004, 5–17.

11. Dastani, M. and van der Torre, L., "Programming BOID Agents: a deliberation language for conflicts between mental attitudes and plans", in: *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04)*, 2004, 706–713.
12. Dastani, M. and van der Torre, L., "What is a normative goal? Towards Goal-based Normative Agent Architectures Regulated Agent-Based Systems", LNAI 2934, Springer, 2004, 210–227.
13. Doyle, J., "Rational Belief Revision", in: Allen, J. F., Fikes, R. and Sandewall, E. (eds.), *KR'91: Principles of Knowledge Representation and Reasoning*, San Mateo, California: Morgan Kaufmann, 1991, 163–174.
14. Galbraith, J. K., *The Affluent Society*, Boston: Houghton Mifflin, 1958.
15. Hansson, S. O., "Logic of Belief Revision", in: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2006.
16. Lindström, S. and Rabinowicz, W., "Epistemic entrenchment with incomparabilities and relational belief revision", in: Fuhrmann, A. and Morreau, M. (eds.), *The Logic of Theory Change*, 93–126.
17. Makinson, D. and van der Torre, L., "Input-output logics", *Journal of Philosophical Logic*, **29**, 2000, 383–408.
18. Makinson, D. and van der Torre, L., "Constraints for input-output logics", *Journal of Philosophical Logic*, **30(2)**, 2001, 155–185.
19. Olsson, E. J., "Lindström and Rabinowicz on relational belief revision", in: T. Ronnow-Rasmussen, J. J., B. Petersson and Egonsson, D. (eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*, 2007.
20. van Riemsdijk, M. B., *Cognitive Agent Programming: A Semantic Approach*, Ph.D. thesis, University of Utrecht, 2006.
21. Smith, D. E., "Choosing Objectives in Over-Subscription Planning.", in: Zilberstein, S., Koehler, J. and Koenig, S. (eds.), *Proceedings of the Fourteenth International Conference on Automated Planning and Scheduling (ICAPS 2004)*, Whistler, British Columbia, Canada: AAAI, 2004, 393–401.